

Michał Bałdyga 184523

Kacper Barański 182613

Data Warehouse Optimization - report

1. Aim of the laboratory

Celem zadania jest przedstawienie zagadnień związanych z różnymi modelami fizycznymi kostki i projektowaniem agregatów.

2. Preliminary Assumption

Size of the database (data warehouse):

Baza danych 18000 wierszy faktów i od kilkudziesięciu do kilku tysięcy wierszy wymiarów.

Testing environment:

System: Windows 11

RAM: 16 GB

Processor: Procesor Intel(R) Core(TM) i5-10300H CPU @ 2.50GHz 2.50 GHz

Microsoft Visual Studio 2022

Microsoft SQL Server Management Studio 2018

3. Testing

W ramach testów wielokrotnie wykonano konkretne zapytania z wykorzystaniem różnych modeli (MOLAP, HOLAP, ROLAP) ze zdefiniowanymi agregacjami, jak i bez zdefiniowanych. Wybrano następujące zapytania:

1. Wybierz 10 dat, w które się odbyło najwięcej wizyt stacjonarnych.[13]

2. Porównaj liczbę e-wizyt dla poszczególnych lekarzy z obecnego oraz poprzedniego miesiąca.[11]

3. . Na jaki rodzaj e-wizyt pacjenci decydują się najczęściej w ostatnim miesiącu.[10]

	MOLAP		ROLAP		HOLAP	
	Aggr	No aggr.	Aggr	No aggr.	Aggr	No aggr.
Querying speed	8ms	10ms	11ms	13ms	9ms	12ms
Zapytanie 1 [13]						
Zapytanie 2 [11]	182ms	226ms	240ms	234ms	238ms	256ms
Zapytanie 3 [10]	7ms	8ms	40ms	49ms	41ms	56ms
Processing time	1144ms	1359ms	799ms	993ms	829ms	783ms
Total size	5.86MB	5.59MB	5.04MB	5.04MB	5.21MB	5.04MB

4. Discussion (comparison of the theory with the obtained results)

Z uzyskanych rezultatów można wywnioskować, że im więcej danych znajduje się w analitycznej bazie danych tym zapytania będą wykonywały się szybciej. Co więcej przechowywane agregacje przyspieszają wykonywanie zapytań, kiedy znajdują się na serwerze OLAP.

Model MOLAP w teorii ma najdłuższy czas procesowania kostki i testy to potwierdziły. Ponadto ten model charakteryzuje się największym rozmiarem, który jest większy dla przypadku z agregacjami i większy dla przypadku bez agregacji co również zgadza się z naszymi wynikami. Większy rozmiar wynika z przechowywania tabeli faktów oraz agregacji w analitycznej bazie danych, jednakże dzięki takiemu sposobowi przechowywania danych czas wykonywania zapytań jest krótszy. Stąd jeżeli zależy nam na prędkości wykonywanych zapytań powinniśmy zastosować właśnie ten model.

Model ROLAP charakteryzuje się najmniejszym rozmiarem bazy analitycznej co jest zgodne z naszymi wynikami. Może on być dobrym rozwiązaniem gdy zależy nam najbardziej na pamięci. Niestety czas wykonywania zapytań jest dłuższy co jest zgodne z teorią. Wynika to z faktu, że na serwerze OLAP znajdują się tylko dane dotyczące kostki, a pozostałe dane jak i zdefiniowane agregacje, znajdują się relacyjnej bazie danych.

Model HOLAP powinien być kompromisem między modelami MOLAP i ROLAP. Agregacje posiada na serwerze analitycznym, a resztę danych w relacyjnej bazie danych. Nasze testy niestety tego nie potwierdziły co może wynikać ze zbyt małej ilości danych. Jednakże można zauważyć, że bez definiowanych agregacji jest on wolniejszy od pozostałych modeli. Rozmiar bazy jest pomiędzy rozmiarami dla pozostałych modeli tak samo jak czas procesowania kostki.

